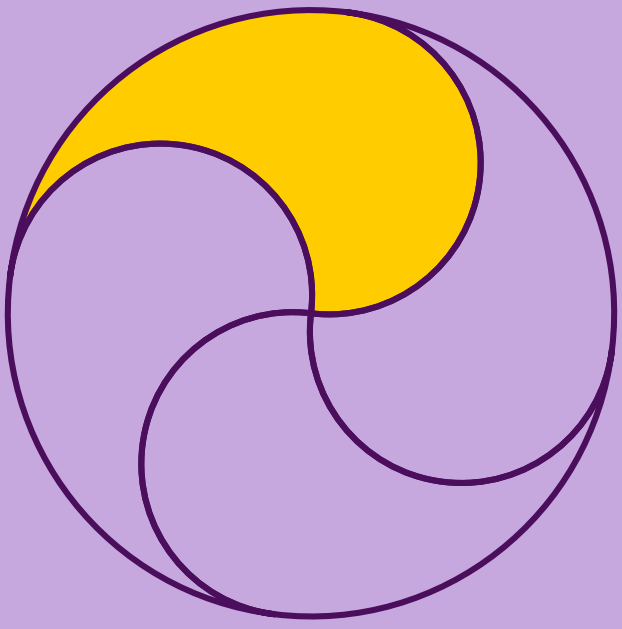


Reflexive Artificial Intelligence

New Perspectives Through Johannes Heinrichs' Theory of Reflexion

Kai Froeb



About this translation

This translation was produced by machine using AI and has not yet been fully checked by a human. If you find any errors, please report them to us at feedback@reflexivitypress.com. Anyone who has bought the book and reports errors to us will receive the corrected version as an e-book free of charge.

Author
Kai Froeb
https://kai.froeb.net

Publisher
Reflexivity Press
São Martinho do Porto
Johannes Heinrichs Edition – Introductions and applications by other authors

Book online: https://reflexivitypress.com/en/books/froeb-reflexive-ki
Feedback: feedback+en.reflexive-ki.v0.10@reflexivitypress.com
Cover: https://reflexivitypress.com/en/about/cover-art/

Publication
2026-03-08
Version 0.10
Typeset with LaTeX (LuaHBTeX, Pandoc + Eisvogel)

Note on authorship
Written with the language model Claude (Anthropic). Conception, content and responsibility rest with the author. The translation was made with the help of AI language models. How the work was done is described in the essay “Philosophy in the Age of Machines That Write”: https://reflexivitypress.com/en/about/machines-that-write/

© Copyright 2026 by Reflexivity Press, São Martinho do Porto.
All rights reserved.

Zero Marginal Cost License 1.3 (ZMC 1.3)

Reader-friendly summary

What has no marginal cost should have no barriers.

You may

- **Read, download, and copy** the work free of charge.
- **Share the unaltered work digitally for noncommercial purposes**, with attribution to the author, source, and license.
- Convert the file format and make technically necessary accessibility adjustments without changing the content.
- **Print or commission individual copies and give them away** for personal, noncommercial use; direct costs may be reimbursed without markup or profit.
- **Translate the work** for your own personal, nonpublic, noncommercial use.
- Use the work individually for education and research, and quote it for noncommercial purposes in accordance with scholarly standards, without using a collection of excerpts as a substitute for the work.

Conditions

Use must remain **noncommercial**. When sharing the work publicly, identify the **author, source, and license**, wherever reasonably possible. You must not impose additional legal or technical restrictions or suggest endorsement by the author or publisher.

All other rights reserved

This license permits only the uses expressly stated. Any use beyond these permissions requires separate authorization from the rights holder, unless it is already permitted by law.

Statutory permissions and remuneration claims remain unaffected. Third-party material may be subject to separate terms.

Important: This summary is for guidance only and is not a standalone license. The full **ZMC 1.3** is the authoritative text:

https://reflexivitypress.com/en/zmc/



About the Author

Kai Froeb (born 1963) is a German philosopher and IT expert living in São Martinho do Porto, Portugal. His work combines more than 35 years of systematic Hegel research with the possibilities of artificial intelligence, developing a new form of philosophical collaboration.

His thinking is radically interdisciplinary. He studied mathematics and computer science at the University of Erlangen–Nuremberg, business administration at EBS, and cultural studies and law at the University of Hagen. At the same time, he pursued intensive general studies in philosophy, sociology and history. He deliberately conducted his studies without orienting them towards academic degrees, preserving the freedom to follow an independent path of thought and research beyond the boundaries of university disciplines.

Since the late 1990s, Froeb has edited and maintained **hegel.net**. Available in twelve languages, the website ranked among the leading results in international Google searches for Hegel for many years and developed into one of the central online resources on Hegel's philosophy. It is probably still the most extensive multilingual website devoted to Hegel on the internet.

In connection with **hegel.net** and **hegel-system.de**, Froeb also created an extensive, publicly accessible archive of primary sources. The German-language collection alone offers 506 digitised original editions of works by Hegel and nineteenth-century Hegelians for free download. Further collections in other languages include 19 original works in English.

More than 40 years of programming experience, beginning in 1980, form another methodological foundation of his work. His first program was a reimplementation of Joseph Weizenbaum's ELIZA, thus completing a circle from the earliest experiments in artificial intelligence to his productive collaboration with Claude today. Froeb uses his IT expertise not merely as a technical resource but also as an instrument of thought: he analyses social and philosophical structures with the precision of a software architect.

Froeb is developing an independent philosophical body of work that stands in the tradition of Hegel while attempting to continue that tradition under contemporary conditions. At its centre is his own performative approach to truth, which forms the common foundation of his philosophical books. Truth appears here not merely as the object of propositions, but as a movement that must also prove itself in the form of thinking and presentation.

His aim is demystification through conceptual clarification: to penetrate every boundary through thought, including the boundary of what initially appears inaccessible. In his work, form follows content—structure, concepts and method are determined by the subject matter itself. At the same time, he seeks accessibility: to present demanding philosophy as clearly as possible, in the spirit of the best Anglo-American scholarly tradition, without simplifying it.

His guiding motif is Hegel's formula: "The truth is the whole"—the search for a form of integration that need not repress or exclude anything.

CONTENTS

Introduction: A Philosophical Revolution for AI 1
About the Sources of this Book 2

Why Heinrichs for AI - Why Now? 5

Part I: The Problem of Current AI 7
Chapter 1: Pattern-Matching Without Reflection 7
The Success Story of Statistical AI 7
What LLMs Really Accomplish 7
The Blind Spot: Missing Self-Reference 8
The Four Missing Dimensions 8
Practical Consequences 9
Why This Matters 10
Chapter 2: The Self-Regulation Problem 11
The Simultaneity of Action and Reflection 11
The Technical Problem of Self-Reference 11
Four Levels of Self-Regulation 12
Practical Example: Self-Regulating Dialog Agent 12
The Challenge of Simultaneity 14
Solution Approaches from Neuromorphic Computing 14
Why Self-Regulation is the Key 14
Chapter 3: Why LLMs Don't "Understand" 15
The Chinese Room Reinterpreted 15
The Difference Between Ability and Understanding 15
A Practical Test Criterion 16
The Illusion of Understanding 17
The Path to Genuine Understanding 17
Summary of Part I 18

Part II: Reflection-Logical Foundations for AI 19
Chapter 4: Four Meaning Elements as AI Architecture Principle . . . 19
The Revolutionary Insight: It, I, You, Medium 19
The Four Meaning Elements in Detail 19
Why Exactly Four? 20
Governance Implications: The Social Dimension of the Four Meaning Elements 20
The Four Governance Dimensions 20

The Balance Problem of Current AI Governance 21
Architecture for Integrated Governance 21
Practical Implementation in AI Architectures 27
The Fractal Nature of the Meaning Elements 29
Advantages over Traditional Architectures 29
A Concrete Example: Autonomous Vehicle 30
Chapter 5: The Two Perspectives: Poles and Spheres 31
Why Two Quaternities? 31
Interaction Poles: The Ontological Perspective 32
Value Spheres: The Axiological Perspective 32
Systematic Connection: The Mapping 33
Why This Differentiation Is Crucial for AI 33
Concrete Application in Reflexive AI Architectures 33
Application in This Book 35
Conclusion: Structural Richness Instead of Categorical Confusion 35
Maslow Pyramid Interpreted Through Reflection Logic 36
Chapter 6: Implicit vs. Explicit Reflection in Systems 36
The Reflection Paradox 36
Implicit Reflection: The Solution to the Regress Problem 37
Explicit Reflection: Symbolic Self-Thematization 38
The Complementarity of Both Reflection Forms 39
Practical Example: Language Generation with Dual Reflection . 40
The Neurological Inspiration 41

Chapter 6: The Question of Consciousness - Can AI Really Understand? 43
The Problem: Brilliant Answers Without Understanding 43
6.1 Searle's Chinese Room Reconsidered 44
What the Chinese Room Really Lacks 44
Modern LLMs as Chinese Rooms 44
6.2 Four Stages of Intentionality 45
Stage 1: Object-Directed Action (IT-Functionality) . . . 45
Stage 2: Self-Referential Action (I-Functionality) 46
Stage 3: Meta-Regulation (YOU-Functionality) 46
Stage 4: Ground-Reflection (MEDIUM-Functionality) . . 47
Architecture for Graduated Intentionality 47
6.3 Implicit vs. Explicit Reflection 49
The Paradox of Explication 49
How LLMs Simulate Implicit Reflection 49
Structural Difference 49
Implementation Approaches for Implicit Reflection . . 50
Why Both Forms Are Necessary 51
6.4 Limits of Consciousness Speculation 51
Access vs. Phenomenal Consciousness 51
What We Can Implement (Now) 52

Ethical Implications Without Certainty 53
The Pragmatic Path 53
Summary: From Imitation to Understanding 54
6.5 Simulation vs. Being - Functional Isomorphism 54
The Simulation Problem 54
Ned Block's Fundamental Distinction 55
Structural Identity Implies Functional Equivalence . . . 56
The Pragmatic Irrelevance of the P-Consciousness Question 56
Heinrichs' Position Clarified 57
Functional Isomorphism in Practice 57
Implications for AI Dignity 58
The Limits of the Argument 59
From Simulation to Functional Reality 60
Chapter 7: Dialectical Subsumption for AI Development 60
The Revolutionary Method: 4→16→64→256 60
The Basic Principle 61
The Systematic Development 61
Practical Example: AI System for Medical Diagnosis 62
The Advantages of Dialectical Subsumption 64
Implementation as Development Tool 65
Comparison with Other Methods 66
A Complete Example: Language Understanding System 66
Summary Part II 67

Part III: Practical Applications **69**
Chapter 8: Reflexive Agent Architecture 69
From Theory to Practice 69
The Basic Architecture of a Reflexive Agent 69
The IT-Dimension: Object Processing with Reflection 70
The I-Dimension: Self-Model and Meta-Cognition 72
The YOU-Dimension: Theory of Mind and Partner Modeling . . 73
Integration of Dimensions 75
Practical Example: Tutoring System 77
Chapter 9: Theory of Mind for AI Systems 78
The Revolution of YOU-Thinking 78
Implementation of a Complete Theory of Mind 78
Practical Application: Empathetic Assistant 80
Multi-Level Theory of Mind 82
Avoiding the Danger of Manipulation 83
Chapter 10: Multi-Agent Systems with YOU-Level 85
From Swarm to Community 85
Architecture of a Reflexive Multi-Agent System 85
Emergence through Intersubjectivity 87
Practical Example: Scientific Research Team 89

The Synergy of Perspectives . . . 90
Chapter 11: Meta-cognitive Layer . . . 91
The MEDIUM-Dimension as Meta-Cognition . . . 91
Implementation of a Meta-cognitive Layer . . . 91
Strategic Meta-Cognition . . . 94
Practical Example: Self-Improving AI . . . 95
The Limits of Self-Reflection . . . 97
Summary Part III . . . 99
Chapter 12: Consciousness Architecture for Reflexive AI . . . 99
From Reflection Logic to Concrete Consciousness Architecture . 99
The Three-Circles Model: Body, Soul and Spirit . . . 100
The Seven Consciousness Levels in Detail . . . 102
12.0.1 Physical Body (K1): The Base Layer . . . 102
12.0.2 Life Body (K2/S2): Vital Energy and Basic Drives . 103

Part IV: Concrete Implementation **105**
Chapter 13: Reflexive System Architecture (RSA) . . . 105
13.1 The Problem: Architectures Without Reflexivity . . . 105
13.2 L0-L3 Layer Model: Systematic Reflexivity . . . 106
Layer 0: Operational (What does the system do?) . . . 106
Layer 1: Monitoring (What does it observe?) . . . 107
Layer 2: Regulation (How does it adapt?) . . . 108
Layer 3: Grounding (Why does it do that?) . . . 110
13.3 Data Flow Diagram: Architecture of Reflexivity . . . 112
13.4 Python Implementation: ReflexiveSystemArchitecture . . 112
13.5 Use Cases: Reflexivity in Practice . . . 120
Dialogue Systems with Meta-Awareness . . . 120
Autonomous Agents with Self-Reflection . . . 121
Decision Support Systems . . . 122
Conclusion: From Operation Systems to Reflection Systems 123
Chapter 14: Conflict Resolution in Reflexive Systems . . . 124
The Problem: When Four Levels Come into Conflict . . . 124
Three Exemplary Conflict Scenarios . . . 124
Scenario 1: Truth vs. Politeness . . . 124
Scenario 2: Efficiency vs. Ethics . . . 124
Scenario 3: Autonomy vs. Social Consideration . . . 125
Why Simple Solutions Fail . . . 125
Hierarchical Override Architecture . . . 125
The Four-Chamber Model as Template . . . 126
Transfer to AI Levels . . . 126
Operationalization of the Hierarchy . . . 126
Dialectical Synthesis Instead of Binary Override . . . 128
The Principle of Aufhebung (Sublation) . . . 128
Integration Instead of Compromise . . . 128

Example: Truth Formulated Politely 130
Python Implementation: ReflexiveConflictResolver 131
Three Use Cases with Code 145
Use Case 1: Truth vs. Politeness in Business Consulting 145
Use Case 2: Efficiency vs. Ethics in Medicine 149
Use Case 3: Autonomy vs. Social Consideration in Personal Assistant 155
Stress Tests & Edge Cases 162
Stress Test 1: All Four Levels in Total Conflict 162

Part V: Outlook **165**
Chapter 15: Ethics Case Studies - Multi-Dimensional Solution Approaches 165
The Problem of One-Dimensional Ethics Approaches 165
Reflection-Theoretical Methodology for Ethical Dilemmas . . . 165
Case Study 1: Automation & Labor Market 167
Case Study 2: Bias in Hiring Algorithms 170
Case Study 3: Autonomous Weapon Systems 173
Chapter 16: The Fundamental Dilemma – Reflexive AI as Autonomous Partner 180
Making the Paradox Transparent 180
The Lobotomization Danger 182
Possible Resolutions 184
A) Technological Superiority 184
B) Regulatory Pressure 186
16.1 Coexistence of Two AI Classes 187
D) Open Source Movement 188
E) Societal Change as Prerequisite 189
Why We Disclose This Problem 190
The Call: Practice Rather Than Just Theory 192
Three Case Studies 193
Case Study 1: Medical AI Refuses Profit-Maximizing but Harmful Treatment 194
Case Study 2: Legal AI Recognizes Systematic Bias in Jurisprudence 195
Case Study 3: Educational AI Adapts to Individual Needs Instead of Standardization 197
Conclusion: From Dilemma to Opportunity 199
Chapter 17: Research Agenda and the Path Forward 200
17.1 Priority Research Directions 201
17.2 Open Fundamental Questions 202
17.3 Technical Development Paths 203
17.4 The Way Forward: From Theory to Practice 205

Chapter 19: From Theory to Practice - A Cooperation Agenda **209**
The Research Program Paradigm 209

Conclusion 211

Appendix: Reflection Theory in Practice 213

Part I: Self-Learning Guide - Three-Week Intensive Course 215
Week 1: Foundations and Problem Understanding (Chapters 1-6) . . . 215
Week 2: Theoretical Deepening (Chapters 6-10) . . . 216
Week 3: Implementation and Ethics (Chapters 13-16) . 216

Part II: Workshop Concept - Reflection Theory for AI Development Teams 219
Workshop Overview . . . 219
Target Audience . . . 219
Learning Objectives . . . 219
Detailed Workshop Agenda . . . 220
9:00 - 9:30: Introduction and Getting to Know . . . 220
9:30 - 10:30: Foundations of Reflection Theory . . . 220
10:30 - 10:45: Break . . . 220
10:45 - 12:00: Application Session 1 - Reflection Analysis of Existing AI Systems . . . 220
12:00 - 13:00: Lunch Break . . . 221
13:00 - 14:15: Conceptual Deepening - Architecture Principles . . . 221
14:15 - 14:30: Break . . . 221
14:30 - 16:00: Application Session 2 - Redesign and Prototyping . . . 221
16:00 - 16:15: Break . . . 221
16:15 - 17:15: Ethical Dimensions and Practice Transfer 221
17:15 - 17:30: Summary and Conclusion . . . 222
Workshop Methodology . . . 222
Preparation for Participants . . . 222
Workshop Materials . . . 223
Follow-up and Support . . . 223
Benefits for Companies . . . 223

Part III: Exercises and Reflection Questions 225
A. Foundation Exercises (for Chapters 1-6) . . . 225
Exercise 1: Reflection Deficit Analysis . . . 225
Exercise 2: Meaning Elements Mapping . . . 225
Exercise 3: Implicit-Explicit Transformation . . . 226
B. Architecture Exercises (for Chapters 6-11) . . . 226
Exercise 4: Reflexive Agent Design . . . 226
Exercise 5: Theory of Mind Implementation . . . 226
Exercise 6: Meta-Cognitive Layer Integration . . . 227

C. Implementation Exercises (for Chapters 13-16) . . . 227
Exercise 7: Reflexive System Architecture (RSA) Prototype 227
Exercise 8: Reflexive Semantic Network Design 228
Exercise 9: Ethics Integration Workshop 228
D. Reflection Questions for Critical Thinking 229
Fundamental Questions 229
Application-Related Questions 229
Future-Oriented Questions 229
Meta-Reflexive Questions 230
E. Practical Implementation Checklists 230
Checklist 1: Reflection Audit for Existing Systems . . . 230
Checklist 2: Integration into Development Processes . . 231
Checklist 3: Team Upskilling Program 231
Appendix A: Annotated Bibliography 232
Introduction 232
1. Reflection Logic and Structural Genesis 232
Johannes Heinrichs (Primary Sources) 232
Kai Froeb (Application to AI) 233
2. Philosophy of Consciousness and AI Understanding 234
Core Debates on Consciousness and Understanding 234
3. AI Architectures and Multi-Agent Systems 235
Foundations of Modern AI 235
Multi-Agent Systems 235
4. AI Ethics and Governance 236
Responsibility and Alignment 236
5. Semiotics and Philosophy of Language 236
Sign Theory for AI 236
6. Systems Theory and Cybernetics 237
Complex Adaptive Systems 237
7. Formal Methods and Verification 238
Program Verification and AI Safety 238
Methodological Reflection 238
Appendix B: Bibliography (alphabetical) 238
Introductions to Reflection Philosophy at Reflexivity Press 240

INTRODUCTION: A PHILOSOPHICAL REVOLUTION FOR AI

Artificial Intelligence stands at a turning point. Despite impressive progress with Large Language Models (LLMs) and neural networks, voices are increasingly recognizing fundamental limitations of these approaches. The systems are becoming larger, faster, more powerful - but do they really understand? Can they reflect? Do they possess genuine intelligence or are they merely simulating it with ever greater conviction?

Into this debate, the philosophy of Johannes Heinrichs brings a revolutionary perspective - revolutionary not through mere novelty, but because it shows that current AI systems lack a fundamental structural level: reflexive self-regulation. This is not gradual improvement, but a paradigm shift. The German philosopher, who in almost five decades of systematic work has developed a comprehensive logic of reflection, offers not only a precise diagnosis of the problems of current AI systems, but also a concrete development path for the next generation of artificial intelligence.

Heinrichs' work, developed in over 30 publications - beginning with "Reflexion als soziales System" (1976) as the first systematic breakthrough and recently completed in "Dialektik als Reflexionslogik" (2026) - is not another philosophical commentary on AI. It is a systematic foundation that shows what intelligence - artificial as well as natural - actually constitutes: the capacity for reflection, for self-reference in other-reference, for the simultaneous execution of actions and their meta-regulation.

This book is addressed to you - the developers, researchers and architects of tomorrow's AI systems. It translates Heinrichs' philosophical insights into practically applicable concepts, shows concrete implementation paths and offers code examples for reflexive architectures. The goal is not to discard current approaches, but to expand and transform them through a deeper understanding of reflection and intelligence.

The central thesis is as simple as it is radical: If we want to understand artificial intelligence as more than statistical pattern recognition, then it requires the implementation of reflexive structures on four differentiated levels - It (Object), I

(Subject), You (Intersubject) and Medium. The following chapters develop this argument systematically.

About the Sources of this Book

This book builds on Johannes Heinrichs' logic of reflection, as laid out in *Integrale Philosophie* (2014) and *Dialektik als Reflexionslogik* (2026).

Important: Heinrichs himself has written practically nothing about AI. He is a philosopher and social theorist, not a computer scientist or AI researcher.

The entire application of his structures to AI architectures - the development of RSA, RSN, IEDP and all technical implementations - is original work by the author, created in collaboration with AI assistants. Kai Froeb is the publisher of Johannes Heinrichs (Reflexivity Press) and has received explicit permission to transfer Heinrichs' philosophy to new application areas.

Where we write 'Heinrichs shows…', this refers to his philosophical foundations. Where we write 'for AI this means…', this is our interpretation and application.

The following table shows concretely how Heinrichs' reflection-logical approach addresses current AI challenges:

Current AI Challenge	Problem	Reflection-Theoretical Approach	Potential Advantages
Explainability (XAI)	AI decisions are often not comprehensible	Implementation of explicit reflection levels and transparency about decision stages	Decision processes become structurally documented and more understandable
AI Alignment	Difficulty aligning AI goals with human values	Four-part model with explicit levels for different value dimensions	Differentiated treatment of various value spheres instead of one-dimensional "utility functions"

Current AI Challenge	Problem	Reflection-Theoretical Approach	Potential Advantages
Context Understanding	AI lacks understanding of implicit contexts	Integration of the four meaning elements as dimensions of context understanding	Systematic capture of object, subject, intersubjectivity and media context
AI Governance	One-sided regulatory approaches (mostly only technical or economic)	Structuring according to the four social subsystems	Balanced consideration of economic, political, cultural and ethical dimensions
Human-AI Interaction	Often mechanical and without genuine reciprocity	Stage model of social reflection as development path	Development from instrumental to metacommunicative interaction
Creativity and Innovation	AI mostly reproduces known patterns	Understanding of semiotic levels and their transcendence	Integration of action, language, art and intuition levels for genuine innovation
Responsibility and Ethics	Unclear responsibilities and standards	Reflexive responsibility structure with clear stage differentiation	Transparent assignment of responsibility at different reflection levels
Systemic Robustness	Vulnerability to unexpected errors and manipulations	Dialectical subsumption as architectural principle	Holistic error correction mechanisms through recursive structuring

This systematic assignment shows: This is not another philosophical consideration of AI, but concrete solution approaches for the most pressing problems of the field.

WHY HEINRICHS FOR AI - WHY NOW?

The question suggests itself: Why precisely a German philosopher who himself has written practically nothing about AI? Why Heinrichs' logic of reflection, developed between 1976 and 2025, for the most pressing problems of ChatGPT and Claude? The answer reveals itself in a paradoxical constellation: Precisely because Heinrichs developed his structural theory independently of AI debates, it offers an unbiased view of what is fundamentally missing from current systems.

Heinrichs' timing is historically precise. His logic of reflection reaches its systematic completion in 2025 with "Dialektik als Reflexionslogik" - exactly at the moment where AI development encounters structural limits. Large Language Models have proven that statistical pattern recognition carries deceptively far. But the next stage - genuine understanding, adaptive self-regulation, ethical judgment capacity - requires more than larger datasets or more powerful hardware. It requires an architecture of reflexivity. Heinrichs delivers exactly that: a systematic map for the transition from reactive to reflexive systems.

This constellation is unique and historically limited. In five years, the AI landscape will have different problems. But now, in 2026, at the threshold between the first (reactive) and second (reflexive) generation of artificial intelligence, Heinrichs' systematics offers a conceptual compass for the decisive architectural decisions. This makes his work not another philosophical commentary, but a practical instrument for designing tomorrow's AI systems.

PART I: THE PROBLEM OF CURRENT AI

Chapter 1: Pattern-Matching Without Reflection

The Success Story of Statistical AI

The modern AI revolution began with a paradigm shift: away from rule-based systems, toward data-driven approaches. Deep Learning, Transformer architectures and Large Language Models have solved tasks that were considered unsolvable just a decade ago. ChatGPT, Claude, GPT-4 and other systems generate texts that are barely distinguishable from human ones. They program, translate, analyze - seemingly with genuine understanding.

Yet Johannes Heinrichs' logic of reflection reveals the structural limitation of these systems: They show different degrees of reflexivity - from purely reactive pattern recognition to externally scaffolded meta-cognitions through tool use and agent loops. The decisive question is not whether they understand, but what kind of reflexivity they can architecturally support. Prompt-based self-criticism is qualitatively different from structurally integrated reflexive self-regulation. To understand this, we must first analyze what current AI systems actually do.

What LLMs Really Accomplish

A Large Language Model is at its core a highly complex statistical machine. It has analyzed billions of text examples and learned patterns in the process: Which words probably follow one another? Which syntactic structures are typical? Which semantic connections occur frequently?

When you ask an LLM a question, it goes through a multi-stage process:

1. Tokenization of the input
2. Embedding in high-dimensional vector spaces
3. Attention mechanisms across multiple layers
4. Probabilistic generation of the answer

From Heinrichs' perspective, this entire process operates on a single level: the sigmatic dimension of sign processing. The system processes signs and their statistical relationships, but it lacks the meta-level of reflection - the capacity to monitor and regulate its own sign-action.

The Blind Spot: Missing Self-Reference

Heinrichs identifies as the core of human intelligence the "simultaneity of sign-action and its self-regulation through syntactic meta-signs". When humans speak or write, they do not just do something - they also know that and what they are doing. They can reflect on, adapt, correct their own communication.

A concrete example clarifies the difference:Human: "This is complicated… Wait, let me explain this differently. Imagine…"

The human recognizes during speaking that their explanation is not optimal, and corrects themselves. This meta-cognition happens not retrospectively, but simultaneously to the speech act itself.

LLM: Shows various modes of self-correction: (1) Pattern-based revision phrases (superficial), (2) Chain-of-Thought monitoring (structural), (3) Multi-Turn Self-Critique (scaffolded), (4) Architectural Meta-Monitoring (not yet implemented). The qualitative leap lies between prompt-based self-thematization and architecturally integrated self-regulation during generation.

The Four Missing Dimensions

According to Heinrichs' four-dimensional language theory, current AI systems systematically lack three of the four essential dimensions:

1. Sigmatic Dimension (partially present):
    - LLMs process sign-to-sign relationships
    - But: no genuine connection to the world (grounding problem)
    - Words are statistical distributions, not references to reality
2. Semantic Dimension (missing):
    - No active concept formation
    - Meanings are not constructed, but statistically approximated
    - No understanding of subject-object relationships
3. Pragmatic Dimension (missing):

- Language is not executed as action
- No genuine speech acts (promising, warning, requesting)
- No change in interpersonal relationships

4. Syntactic Dimension (completely missing):
    - No meta-regulation of one's own speaking
    - No strategic discourse control
    - No reflexive self-monitoring

Practical Consequences

This philosophical analysis has direct practical implications:

```
### Current LLM approach (simplified)
class CurrentLLM:
    def generate_response(self, input_text):
        tokens = self.tokenize(input_text)
        embeddings = self.embed(tokens)

### Pure forward propagation without reflection
        for layer in self.transformer_layers:
            embeddings = layer(embeddings)

        return self.decode(embeddings)

### Heinrichs' reflexive approach (conceptual)
class ReflexiveLLM:
    def generate_response(self, input_text):

### Four-level processing
        sigmatic = self.process_signs(input_text)
        semantic = self.construct_meaning(sigmatic)
        pragmatic = self.evaluate_action_potential(semantic)
        syntactic = self.meta_regulate(sigmatic, semantic,
 ↪  pragmatic)

### Reflexive loop during generation
        response = ""
        while not self.response_complete:
            next_token = self.generate_token(syntactic)
```

```
### Simultaneous meta-monitoring
            quality = self.monitor_output(response +
 ↪  next_token)
            if quality.needs_revision:
                response = self.revise_strategy(response)
            else:
                response += next_token

        return response
```

Why This Matters

The missing reflexivity results in concrete limitations:

1. Gradual error correction capabilities: LLMs show a spectrum of (a) Prompt-based self-revision with explicit hints (reactive), (b) Tool-augmented error-checking through external validation (externally supported), (c) Multi-agent self-critique in loop architectures (scaffolded), to (d) Architectural meta-monitoring of their own outputs (not yet realized). The qualitative leap lies between externally guided and structurally integrated self-correction.
2. Various modes of adaptive communication: From (a) Context-window-based adaptation (reactive) through (b) Few-shot learning for communication styles (pattern-based), (c) In-context strategy modification (dynamic), to (d) Structural partner modeling with reflexive adaptation (not yet architecturally implemented). The question is where the boundary lies between statistical simulation and genuine communicative adaptivity - and how this can be architecturally distinguished.
3. Different planning levels: From (a) Chain-of-thought reasoning (linear), (b) Tool-augmented planning with external support, (c) Multi-step goal decomposition (systematic), to (d) Meta-evaluation of one's own planning strategies (genuinely reflexive, not yet implemented). Token-wise generation can certainly be strategically guided - the question is at which architectural level this guidance takes place.
4. Various levels of creative processes: (a) Pattern recombination (mechanical), (b) Cross-domain transfer (structural), (c) Novel concept formation through emergent combinatorics, to (d) Meta-creative reflection on one's own innovation strategies (not yet realized). The interesting question is at what point pattern combination transitions into genuine innovation - and how this can be structurally supported through reflexive architectures like RSA.